June 19, 2015

Attachment 2
Reference Data Selection Process and Sediment Quality Triad Methodology for use in the LPRSA 17-Mile RI/FS

Reference Data Selection Process (Censoring of Reference Data Sets)

The reference data provide a measure of regional conditions against which to compare the biological response measures in Lower Passaic River (LPR) stations. In order to evaluate the effect of chemical contaminant exposure on the macroinvertebrate endpoint in the LPR, the reference datasets should be representative of all regional stressors that could affect macroinvertebrate community structure and population survival/growth metrics absent this stressor category. Accordingly, the reference datasets need to be censored by removing all stations that exhibit elevated contaminant concentrations. Each of the four reference datasets -- Above Dundee Dam (all data collected, not just the area closest to the dam), Jamaica Bay, Mullica River (include freshwater data collected as part of Berry's Creek investigation), Mullica River/Great Bay -- should have the following two methods employed to censor them:

A. Apply applicability criteria that include acceptable laboratory test survival and lack of sediment benchmark exceedances. Reference stations should be identified using the following criteria:

   1. laboratory bioassay results >= 80% of negative control (estuarine); >=75% of negative control (freshwater)
   2. comparison to sediment quality criteria
      a. Marine/Estuarine
         i. no exceedances of Effect Range- Median (ER-M) benchmarks,
         ii. no more than 3 exceedances of Effect Range – Low (ER-L) [see Weisberg et al 1998]
      b. Freshwater [see MacDonald et al 2000] – Mean Probable effect concentration quotient (PECq) <0.5

B. Interquartile Range, where the Q3-Q1 x 1.5 provides the value for determining if a result is an outlier.

Sediment Quality Triad Methodology

A. **Benthic Community:** There should be two comparisons made for the benthic community metrics (abundance, taxa richness, Shannon-Weiner, Pielou's, Swartz, and Hilsenhoff (freshwater)) once the reference datasets are appropriately censored. The first, as is done in the draft BERA, is to compare the means using a Mann-Whitney test. The second is to compare each station to the 5^{th} percentile of the reference dataset. If the individual station is less than (or greater depending upon the metric) the reference value, it is considered impacted for that measure.

B. **Benthic Toxicity:** Sediment toxicity tests should be compared to the negative control and the appropriate reference locations.

1. Statistical comparison to the negative control should use one of the following approaches depending on the dataset distribution:

   a. one-tailed t-test (with equal or unequal variances)(Zar 1996).
   b. The non-parametric test, which is the same one-tailed equal variance t-test but performed on the "rankit scores" (aka normal scores or normalized ranks). Rankit scores are used in parametric analyses to provide non-parametric alternatives (Conover 1980; Clarke and Brandon 1996).

   The most appropriate test can be determined based on the outcome of a test for normality (Shapiro-Wilk's test on the residuals, alpha=0.05) and Levene's test for equality of variances (alpha=0.05) (Conover 1980; Clarke and Brandon 1996). Multiple comparison tests (e.g., ANOVA or Dunnett's) should not be used.

2. The 5^{th} (and/or the 95^{th}, potentially, in the case of the abundance community metric) percentile estimates for biological response parameters (laboratory survival and benthic community metrics) from the appropriately censored reference data sets should be compared to the appropriate values at each SQT station.

   Additionally, the uncertainty section should clearly state that the LPR *Ampelisca* test did not follow the same ASTM/EPA protocol that was used for the reference data sets, and that it likely underestimates toxicity compared to reference.

Sediment Chemistry: Sediment chemistry is an important component of the sediment quality triad. The draft BERA compared sediment chemistry data to NJDEP Sediment Quality Criteria, which are used for screening purposes. The comparison criteria that should be used are T20 and T50 values and the chemistry for each individual sample should be evaluated (Field et al 2002). An additional statistical evaluation, that does not receive a scoring value as part of the SQT, should be conducted. As discussed on a previous call, a multivariate analysis should be conducted, possibly in conjunction with a multiple regression analysis. This would replace the bivariate Spearman Rank Correlation analysis. Two methods that were identified by the CPG were principal component analysis and cluster analysis. EPA would need to review the approach being implemented to ensure that the methodology and parameters are acceptable. A third line of evidence that was presented in the BERA, AVS-SEM, should also be retained.

SQT Weighting: There are three categories for the sediment quality triad (Benthic Community, Sediment Toxicity and Sediment Chemistry) with each category having multiple metrics evaluated. Each category should have equal weighting in the analysis so that no one category can result in an impacted or not impacted result. A value of 1 is assigned for each category and the values for each metric are divided equally (see table below). For each station, an SQT score will be calculated and compared to the classification system listed below.

SQT Weighting Table

Category	Metric	Value
Benthic Community (est/fw)	Abundance	0.2/0.16
	Richness	0.2/0.16
	Diversity	0.2/0.16
	Evenness	0.2/0.16
	Dominance	0.2/0.16
	Tolerance of Environmental Stress	na/0.16
Sediment toxicity	*A. abdita* survival – estuarine	0.33
	H. azteca survival – estuarine	0.33
	H. azteca biomass – estuarine	0.33
	C. dilutus survival – freshwater	0.25
	C. dilutus biomass - freshwater	0.25
	H. azteca survival - freshwater	0.25
	H. azteca biomass - freshwater	0.25
Sediment chemistry	Comparison to T20 and T50 values	0 = no exceedance 0.5 = (low) exceedance of T20 1 = (high) exceedance of T50
Classification system	No impact 0 - 0.75 Low impact 0.75 - 1.5 Medium impact 1.5-2.25 High impact 2.25- 3.0	

References

Clarke, Joan U. and Dennis L. Brandon. 1996. Applications Guide for Statistical Analyses in Dredged Sediment Evaluations. U.S. Army Corps of Engineers, Waterways Experiment Station. Technical Report D-96-January 1996.

Conover, W.J. 1980. Practical Nonparametric Statistics, 2nd ed. John Wiley & Sons, New York.

Weisberg, S.B., J.A. Ranasinghe, J.S. O'Connor, D.A. Adams. 1998. A benthic index of biotic integrity (B-IBI) for the New York/New Jersey Harbor. Appendix C in Adams, D.A., O'Connor, J.S., Weisberg, S.B. 1998. Sediment quality of the NY/NJ Harbor System. EPA/902-R-98-001. U.S. Environmental Protection Agency, Region 2, Edison, NJ.

Field, J.L., D.D. MacDonald, S.B. Norton, C.G. Ingersoll, C.G. Severn, D. Smorong and R.Lindskoog. 2002. Predicting Amphipod Toxicity from Sediment Chemistry Using Logistic Regression Models. Environmental Toxicology and Chemistry, Vol. 21, No. 9. Pp. 1993-2005.

MacDonald, D.D., C.G. Ingersoll and T.A. Berger. 2000. Development and Evaluation of Consensus-Based Sediment Quality Guidelines for Freshwater Ecosystems. Arch. Environ. Contam. Toxicol. 39, 20-31.

Zar, Jerrold H. 1996. Biostatistical Analysis, Third Edition. Prentice Hall, New Jersey.